

METHOD FOR DETECTING MICROORGANISMS

The present invention relates to a method for detecting the elements constituting microorganism flora, at least some of the elements of which have an operon in common, characterized in that the elements of said flora are identified by studying the intergenic sequence of said operon.

10 The human digestive system harbors a considerable number of microorganisms which constitute a microbial flora of extreme complexity. Although these bacteria are distributed throughout the digestive tract, the colon contains most of the flora, both from a  
15 quantitative and a qualitative point of view. It is estimated that the colonic flora of an individual consists of  $10^{13}$  to  $10^{15}$  bacteria, mostly anaerobic bacteria, represented by at least 400 species belonging to approximately 30 different genera. These bacteria  
20 colonize the various stages of the colon in a relatively heterogeneous manner. A "fermentation flora" in the cecum and also a "putrefaction flora" in the left colon are conventionally described. Moreover, a residence flora is distinguished from a passing flora.  
25 This residence flora is itself divided up into dominant

flora and subdominant flora. The dominant bacteria, essentially anaerobic bacteria, are mainly represented by the *Bacteroides* genus, gram-negative bacilli, but also by the *Bifidobacterium*, *Lactobacillus* or *Clostridium* genera, gram-positive bacilli. The subdominant flora contain aero-anaerobic and microaerophilic bacteria. The presence of enterobacteria and of streptococci is especially noted. Diet, infections, intake of pre- and probiotics and also treatments with antibiotics are all liable to cause drastic modifications in the composition of the colonic flora. Since these variations have a direct impact on health, it is important to be aware of them and to understand them, both in order to avoid them and in order to trigger them for therapeutic purposes. To date, the study of colonic bacteria has made it possible to characterize approximately 200 species. However, this type of investigation comes up against many problems essentially associated with the laboriousness of the techniques. The results obtained and also the conclusions which ensue therefrom are still sketchy.

Bacterial identification is carried out according to various methods which employ either the demonstration of specific phenotypical or biochemical character-

istics, or the use and recognition of specific heterologous regions present on the genome.

The beginnings of an identification may be carried out  
5 by describing the morphology of the organism studied  
and by searching, for example, for the presence of  
endospores, of sheaths, of cysts, of buds, of fruiting  
bodies, etc. The form of the colonies, the  
pigmentation, and the origin of the sample are also  
10 useful pieces of information. Preliminary studies may  
also be carried out using dyes specific for the  
capsule, the flagella, the granules, the wall, etc.  
However, a more thorough and precise identification  
necessarily involves techniques comprising isolation on  
15 selected media. Such media are developed or improved in  
order to increase the specificity of the selection.  
However, it is difficult to completely exclude the  
presence of possible contaminants which may then  
interfere in the recognition tests subsequently  
20 employed.

These tests make use of the biochemical characteristics  
specific for a species. Multitest systems exist. They  
are identification galleries which are provided in the  
25 form of kits, of microplates and of strips, the use of  
which is sometimes automated. However, there exists a

certain degree of heterogeneity of chromosomal or plasmid origin within many species. Thus, one or more characteristics will be absent in the identification. Thus, most of the time, only the probability of belonging to a species will be given. A similarity of 80% or more with a reference bacterium will be considered as acceptable. Single-test systems have also been developed. They use synthetic fluorescent substrates for revealing the presence of an enzyme specific for a microorganism. They enable a rapid analysis, but are limited in their use. Specifically, a specific test must be developed for each species.

Immunoassays using poly- or monoclonal antibodies have also been developed. Besides the obvious limitation inherent in polyclonal antibodies, these assays are mainly used to characterize serotypes, but are rarely used to identify species. Although commonly used in hospital diagnosis, they remain relatively unused in bacteriology. As regards the production of monoclonal antibodies, it remains long, laborious and expensive work. They are, for example, directed against lipopolysaccharides, the membrane, the pili, etc., but are, however, rarely specific for a species.

The development of molecular biology techniques has made it possible to develop novel identification assays. These techniques are based on hybridization reactions or polymerase chain amplification reactions  
5 (PCR).

Genome/genome hybridization should be greater than or equal to 70% to identify an unknown bacterial species as being the known reference bacterium whose genomic  
10 DNA is used to perform the diagnosis. Other assays involve heterologous regions on the DNA specific for a species. The hybridization technique can thus be used, which consists in depositing the product to be analyzed on a nylon or nitrocellulose membrane and then in  
15 incubating with a specific labeled probe (cold probe or hot probe) {1}.

It is also possible to use specific primers making it possible to amplify a fragment of a given size by the  
20 PCR technique {2, 3}. In this case, the amplicates obtained by PCR can themselves be analyzed by other techniques, such as RFLP (restriction fragment length polymorphism) {4} or TGGE (temperature gradient gel electrophoresis) {5}, refining the diagnosis.

Despite their large capacity for discrimination, these techniques remain limited since they make it possible to analyze only one species at a time, or else consist in isolating a mixture of species which cannot then be identified without knowing exactly the pattern of analysis of the amplicates by a given technique on a given biotope.

The development of DNA chips (biochips) makes it possible to envision a rapid diagnosis relating to several hundreds of species. This technique consists in placing on a surface area of a few millimeters squared several hundreds of DNA sequences specific for a given organism. These probes are hybridized with DNA fragments, generally obtained by RT-PCR. The possible hybridization of said fragments is then observed, and indicates the presence or absence of the gene expressed, or of the organism studied.

The nucleic acid targets studied over the last few years are essentially 16S ribosomal RNA (with more than 7000 available sequences) {1, 2, 4, 5}, the region separating the 16S and 23S genetic loci {3} and the elongation factors {6, 7}. Thus, using 16S RNAs as a basis, it has been possible to detect several new species, belonging to the *Bacteroides* and *Clostridium*

groups, but not to the *Bifidobacterium* group {8}.  
Moreover, and although 16S RNAs make it possible to  
detect and to identify many bacteria down to the  
species, they are incapable of discriminating between  
5 the various species of staphylococci {2, 3}. Thus, a  
variable region of the gene encoding HSP 60 has been  
proposed for studying the microorganisms of the  
intestinal flora {9}.

10 A subject of the present invention is a method for  
detecting and identifying the elements constituting a  
microorganism flora, in particular the intestinal  
flora, according to which a target which is even more  
discriminatory and universal than those already studied  
15 is detected and studied.

The method according to the present invention comprises  
characterizing the sequences of this target for the  
organisms present in the microbial flora studied, and  
20 thus makes it possible to design a diagnostic test.

Thus, the target studied in the method of the invention  
exhibits strong interspecies heterogeneity, which  
allows discrimination between the microorganisms.

The present invention therefore relates to a method for detecting the elements constituting a microorganism flora, at least some of the elements of which have an operon in common, characterized in that:

5

a) the genomic DNA of said flora or the mRNAs is (are) prepared,

10 b) at least some of the noncoding intergenic sequences located in the operon conserved in at least some of the elements of the flora are amplified, and

15 c) the various intergenic sequences amplified are identified in order to determine the elements of said flora.

Specifically, surprisingly, it has been noted that the intergenic regions, in the operons conserved between  
20 various species, exhibit a certain heterogeneity, whereas the coding regions which flank said regions in the 5' and/or in the 3' position are generally very conserved. It is possible that this fact is due to a relatively weak selection pressure on the noncoding  
25 regions in the course of evolution {10}.

The amplification is preferably carried out by polymerase chain reaction (PCR), but other methods (PCR-like) may be employed, using a pair of primers having nucleotide sequences for implementing the method  
5 according to the invention.

The term "PCR-like" is intended to denote all methods using direct or indirect reproductions of nucleic acid sequences, or else in which the labeling systems have  
10 been amplified; these techniques are, of course, known. In general, it involves amplification of the DNA with a polymerase; when the sample of origin is an RNA, a reverse transcription should be carried out beforehand. A very large number of methods currently exist for this  
15 amplification, such as, for example, the SDA (Strand Displacement Amplification) technique, the TAS (Transcription-based Amplification System) technique, the 3SR (Self-Sustained Sequence Replication) technique, the NASBA (Nucleic Acid Sequence Based  
20 Amplification) technique, the TMA (Transcription Mediated Amplification) technique, the LCR (Ligase Chain Reaction) technique, the RCR (Repair Chain Reaction) technique, the CPR (Cycling Probe Reaction) technique, or the Q-beta-replicase amplification  
25 technique. Some of these techniques have since been improved.

The analysis of the amplified sequences is advantageously carried out on a DNA kit comprising sequences complementary to the sequences liable to be amplified from the elements of said flora. Knowledge of  
5 the microorganisms which may be present in the biological sample studied is therefore important in order to choose the DNA chip to be analyzed. Thus, it is necessary for the DNA chip to have, at its surface,  
10 probes specific for each of the organisms intended to be studied. Such a DNA chip is also a subject of the invention.

Thus, the present invention relates most particularly  
15 to a DNA chip comprising, at its surface, a plurality of oligonucleotides complementary to the intergenic sequences of the various operons conserved between the species. The term "DNA chip" is intended to mean a solid support to which are attached nucleic acid  
20 fragments under conditions which allow hybridization thereof with the complementary oligonucleotides, and detection of the hybrids thus formed. Thus, a DNA chip according to the invention also relates to the membranes as used to perform Southern blotting.

The oligonucleotides attached to the DNA chip according to the present invention are so attached by any conventional method known to those skilled in the art, and are approximately 50 bases long. It is understood  
5 that the oligonucleotides considered may also be shorter or longer. Thus, it is within the scope of those skilled in the art to determine the length of the oligonucleotides attached to the chip according to the invention, for each sequence.

10

Preferably, the oligonucleotides attached to the DNA chip are chosen such that their sequence comprises a part of the hypervariable region identified according to the present invention. The oligonucleotides attached  
15 to the chip according to the invention may also contain sequences corresponding to the sequences variable to a lesser degree, located at or close to the end of the operon genes.

20 In a particular and preferred implementation of the invention, the DNA chip according to the invention has a plurality (a number greater than or equal to 2, preferably 3, more preferably 5, most preferably 10) of oligonucleotides greater than 40 bases long.  
25 Preferably, said oligonucleotides comprise a fragment of at least 20, preferably 40 or 50, more preferably

75, most preferably 100 consecutive bases, of the sequences SEQ ID No. 63 to SEQ ID No. 138 and SEQ ID No. 140 to SEQ ID No. 189, corresponding to the intergenic sequences of various species (rpoBC for SEQ ID No. 63 to SEQ ID No. 138, GroESL for SEQ ID No. 140 to SEQ ID No. 189).

Thus, demonstrating the possible hybridizations of the amplified sequences makes it possible to identify the elements present in the microbial flora studied.

An operon which is particularly suitable for implementing the method according to the invention is the bacterial rpoBC operon. This bacterial operon contains coding sequences which are relatively homologous between genera. It is therefore possible to determine degenerate primers for amplifying a region which is heterologous between species and which corresponds to the transcribed intergenic region (IGR). In bacteria, the rpoBC operon encodes the beta and beta prime subunits of DNA-directed RNA polymerase, just like the homologous genes, which may or may not be conserved in the form of an operon, in mitochondria and other eukaryotic organelles (chloroplasts), and just like nuclear eukaryotic RNA polymerase II (which synthesizes the messenger RNAs). The study of this

operon makes it possible not only to detect the bacteria, but also other eukaryotic microorganisms (yeast, protozoa, or others).

5 The method according to the invention is thus carried out using degenerate primers located in the coding sequences of the operons, in particular at least one primer chosen from the sequences SEQ ID No. 1 to SEQ ID No. 31, themselves a subject of the invention. The RNA  
10 polymerase proteins are in fact extremely conserved according to the species, which makes it possible to find amino acid sequences which align with one another, and thus to choose degenerate oligonucleotides for amplifying the intergenic sequences.

15

The pairs of primers described by the sequences: (a sequence chosen from the sequences SEQ ID No. 1 to SEQ ID No. 8)/(a sequence chosen from the sequences SEQ ID No. 9 to SEQ ID No. 11) are used to perform a first  
20 amplification of the intergenic, IGR, of the bacteria. A second, more specific, amplification can then be carried out using pairs of primers which hybridize within the first amplified region, and which are described by the sequences: (a sequence chosen from the  
25 sequences SEQ ID No. 12 to SEQ ID No. 15)/(a sequence

chosen from the sequence SEQ ID No. 16 to SEQ ID No. 31).

SEQ ID N°1	GGN GAY AAR YTN GCN GGN AGN CAY GG
SEQ ID N°2	GGN GAY AAR YTN GCN GGN CGN CAY GG
SEQ ID N°3	GGN GAY AAR YTN GCN AAY AGN CAY GG
SEQ ID N°4	GGN GAY AAR YTN GCN AAY CGN CAY GG
SEQ ID N°5	GGN GAY AAR ATG GCN GGN MGN CAY GG
SEQ ID N°6	GGN GAY AAR TTY GCN TCN MGN CAY GG
SEQ ID N°7	GGN GAY AAR TTY GCN AGY MGN CAY GG
SEQ ID N°8	GGN GAY AAR TTY GCN ACN MGN CAY GG
SEQ ID N°9	AAY GCN GAY TTY GAY GGN GAY CAR AT
SEQ ID N°10	AAY GCN GAY TTY GAY GGN CAR ATG GC
SEQ ID N°11	AAY GCN GAY TTY GAY GGN GAY GAR AT
SEQ ID N°12	GGN GGN CAR MGN TTY GGN GAR ATG GA
SEQ ID N°13	GGN GGN CAY GGN TTY GGN GAR ATG GA
SEQ ID N°14	GGN GGN CAR WSN TTY GGN GAR ATG GA
SEQ ID N°15	GGN GGN NTN MGN TTY GGN GAR ATG GA
SEQ ID N°16	GGN AAR CGN GTN GAY TAY TCN GGN MG
SEQ ID N°17	GGN AAR CGN GTN GAY TAY AGN GGN MG

SEQ ID N°18	GGN AAR AGN GTN GAY TAY TCN GGN MG
SEQ ID N°19	GGN AAR AGN GTN GAY TAY AGN GGN MG
SEQ ID N°20	GGN AAR CGN GGN GAY TAY TCN GTN MG
SEQ ID N°21	GGN AAR CGN GGN GAY TAY AGN GTN MG
SEQ ID N°22	GGN AAR AGN GGN GAY TAY TCN GTN MG
SEQ ID N°23	GGN AAR AGN GGN GAY TAY AGN GTN MG
SEQ ID N°24	GGN AAR CGN GTN GAY TTY TCN GGN MG
SEQ ID N°25	GGN AAR CGN GTN GAY TTY AGN GGN MG
SEQ ID N°26	GGN AAR AGN GTN GAY TTY TCN GGN MG
SEQ ID N°27	GGN AAR AGN GTN GAY TTY AGN GGN MG
SEQ ID N°28	GGN AAR CGN GTN GAY TTY TCN GCN MG
SEQ ID N°29	GGN AAR CGN GTN GAY TTY AGN GCN MG
SEQ ID N°30	GGN AAR AGN GTN GAY TTY TCN GCN MG
SEQ ID N°31	GGN AAR AGN GTN GAY TTY AGN GCN MG

5

The pairs of primers described by the sequences: (a sequence chosen from the sequence SEQ ID No. 53 to SEQ ID No. 54 are used to amplify a the intergenic region, IGR, of the bacteria.

FO	SEQ ID N°53	GGN GGN CAN NSN TTY GGN GAR ATG GA
RP	SEQ ID N°54	AAY GCN GAY TTY GAY GGN GAY SAR AT
FO	SEQ ID N°55	GGN GGN CAR MGN TTY GGN GAR ATG GA
	SEQ ID N°56	GGN GGN CAY GGN TTY GGN GAR ATG GA
	SEQ ID N°57	GGN GGN CAR WSN TTY GGN GAR ATG GA
	SEQ ID N°58	GGN GGN NTN MGN TTY GGN GAR ATG GA
RP	SEQ ID N°59	AAY GCN GAY TTY GAY GGN GAY CAR AT
	SEQ ID N°60	AAY GCN GAY TTY GAY GGN CAR ATG GC
	SEQ ID N°61	AAY GCN GAY TTY GAY GGN GAY GAR AT

These primers were designed based on the study of the  
 5 degeneracy of conserved protein motifs corresponding to  
 rpoB and/or encoded by the rpoB gene:

beta 2 I :

coryneb/bif/actinom/camp/pseudom/salmon/esch/vibrio  
 /clos/bact/hel/citrob/prot/haf/yers/  
 /past/actinob/aer

SEQ ID N°55 GGN GGN CAR MGN TTY GGN GAR ATG GA  
 (8 deg)

beta 2 ii : bacillus

SEQ ID N°56 GGN GGN CAY GGN TTY GGN GAR ATG GA  
 (7 deg)

beta 2 iii : helicobacter mustelae

SEQ ID N°57 GGN GGN CAR WSN TTY GGN GAR ATG GA  
 (8 deg)

beta 2 iv : archae (methano)

SEQ ID N°58 GGN GGN NTN MGN TTY GGN GAR ATG GA  
 (9 deg)

FO : 2 I/II/III : GGN GGN CAN NSN TTY GGN GAR ATG GA  
 (SEQ ID N° 53)

For the reverse sequences, determined based on the degeneracy of conserved protein motifs corresponding to rpoC and/or encoded by the rpoC gene

beta p 2 i :  
coryneb/bif/actinom/bac/camp/pseudom/salmon/esch/vi  
brio/clos/bact/hel/citrob/prot  
/haf/yers/past/actinob/aer/staph/lactob/enteroc/lactoc  
SEQ ID N° 59 AAY GCN GAY TTY GAY GGN GAY CAR AT (8  
deg)  
beta p 2 ii : archae (methano)  
SEQ ID N° 61 AAY GCN GAY TTY GAY GGN GAY GAR AT (8  
deg)  
beta p 2 iii : streptoc  
SEQ ID N° 60 AAY GCN GAY TTY GAY GGN CAR ATG GC (7  
deg)  
  
RP : P 2 i/ii : AAY GCN GAY TTY GAY GGN GAY SAR AT  
(SEQ ID N° 54)  
« REVERSE » ATY TSR TCN CCR TCR AAR TCN GCR TT  
(SEQ ID N° 62)

5

These primers are also part of the invention.

A subject of the invention is also the genomic  
10 sequences of microorganisms which may be amplified by  
the primers according to the invention, in particular  
the pairs of primers: (a sequence chosen from the  
sequences SEQ ID No. 1 to SEQ ID No.8)/(a sequence  
chosen from sequences SEQ ID No. 9 to SEQ ID No. 11),  
15 and the pairs of primers: (a sequence chosen from the  
sequences SEQ ID No. 12 to SEQ ID No. 15)/(a sequence  
chosen from the sequences SEQ ID No. 16 to SEQ ID No.  
31). Amplification with pairs of primers: (a sequence

chosen from the sequences SEQ ID No. 53, SEQ ID No. 55 to SEQ ID No. 58)/(a sequence chosen from the sequences SEQ ID No. 54, SEQ ID No. 59 to SEQ ID No. 61) is also envisioned.

5

Thus, a subject of the invention is also in particular a sequence from SEQ ID No. 63 to SEQ ID No. 138, which correspond to the hypervariable intergenic regions of the rpoB operon or various organisms. A subject of the invention is also a fragment of a minimum of 20 bases, preferably 30 bases, more preferably 50 bases, even more preferably 75 bases, most preferably 100 bases of one of the sequences SEQ ID No. 63 to SEQ ID No. 138, or the sequences complementary thereto, it being possible for said fragment to be used to define organism-specific primers, or for the identification of organisms, in particular by hybridization.

Thus, the DNA chip according to the invention preferably has, at its surface, a plurality of oligonucleotides (a minimum of two) comprising fragments chosen from the fragments of the sequences SEQ ID No. 63 to SEQ ID No. 138 defined above, thus allowing the identification of microorganisms. The length of these oligonucleotides can be determined by those skilled in the art, as a function of the

hybridization conditions which they intend to use. Oligonucleotides approximately 50 bases long are thus envisioned.

5 Another operon which is particularly suitable for implementing the method according to the invention is the bacterial GroESL operon. This bacterial operon is bicistronic and contains coding sequences which are relatively homologous between genera. It is therefore  
10 also possible to determine degenerate primers to amplify a region which is heterologous between species and which corresponds to the transcribed intergenic region (IGR). In bacteria, the GroESL operon encodes the HSP10 and HSP60 proteins (heat shock proteins of 10  
15 and 60 kDa respectively), just like the homologous genes, which may or may not be conserved in the form of an operon, in mitochondria and other eukaryotic organelles (chloroplasts). The study of this operon makes it possible not only to detect bacteria, but also  
20 other eukaryotic microorganisms (yeasts, protozoa, or others).

The method according to the invention is thus carried out using degenerate primers located in the coding  
25 sequences of the operons. The HSP proteins are in fact extremely conserved according to species, which makes

it possible to find amino acid sequences which align with one another, and thus to choose degenerate oligonucleotides to amplify the intergenic, promoter or terminator sequences.

5

Preferably, the primers described by the sequences SEQ ID No. 32 and SEQ ID No. 33 are used to amplify the intergenic region, IGR, of *E. coli* and of Enterobacteriaceae.

10

ENT-BDEG:

CTGGAYGTKAARRTNGGYGAYATYGT

(SEQ ID N° 32)

ENT-ADEG:

ANNACNGTNGCRGTRGTGGTIRCCGTC

(SEQ ID N° 33)

Other degenerate primers can also be used to implement the protocol according to the invention, in particular  
15 any primer chosen from the sequences SEQ ID No. 34 to SEQ ID No. 52.

UNI-ADEG 1 :  
GGNGAYGGNACNACNACNGCNACNNT (SEQ ID N° 34)

UNI-ADEG 2 :  
GGNGAYGGNACNACNACNTGNTCNNT (SEQ ID N° 35)

ENT-BNEW :  
AANMTTCGTCCNYTRCANGAYCGNGT (SEQ ID N° 36)

CLO-BNEW2 :  
ATNARRCCAYTWGGWGAYMGNGTWGT (SEQ ID N° 37)

BIF-BNEW :  
AARCCRCTCGAGGACMRNRTNSTSGT (SEQ ID N° 38)

UNI-A3 :  
GGNGAYGGNACNAANACNGCNACNNT (SEQ ID N° 39)

BIF-BNEW2 :  
ATCAAGCCNCTMGRRGACMRSRTNST (SEQ ID N° 40)

HEL-BNEW :  
NTNCANCCNTTNGGNGANAGNGTNTT (SEQ ID N° 41)

CAM-BNEW :  
NTNCANCCNTTNGGNAANCGNGTNCT (SEQ ID N° 42)

BACT-BNEW :  
NTNAANCCNTTNGCNGANCGNGTNCT (SEQ ID N° 43)

CHLA-BNEW :  
NTNAANCCNTTNGGNGANAGNATNTT (SEQ ID N° 44)

MYCP-BNEW :

NTNAAACCNNTNGGNAANCGNGTNAT	(SEQ ID N° 45)
STA-BNEW :	
NTNAAACCNNTNGGNAANCGNGTNAT	(SEQ ID N° 46)
LACC-BNEW :	
TTGAAACCNNTTAGNGRAYCGYGTRST	(SEQ ID N° 47)
LACB-BNEW :	
TTAMARCCAWTMGGNGATCGNGTNRT	(SEQ ID N° 48)
CLO-BNEW3 :	
ATNANACCANTNGGNGACAGNGTNGT	(SEQ ID N° 49)
ENT-BNEW2 :	
NTNCGNCCNTTNCANGANCGNGTNAT	(SEQ ID N° 50)
LEG-BNEW :	
NTNCGNCCNTTNCANGANCGNGTNGT	(SEQ ID N° 51)
AER-BNEW :	
NTNCGNCCNCTNCANGANCGNGTNAT	(SEQ ID N° 52)
LACB-BNEW2 :	
MARCCNNTNGGNGAYMGNGTNATNGT	(SEQ ID N° 139)

These primers are also subjects of the present invention. Preferably, the detection of a microorganism is carried out using a pair of primers SEQ ID No. 32/ 5 SEQ ID No. 33, or (SEQ ID No. 34, SEQ ID No. 35 or SEQ ID No. 39)/(a sequence chosen from the sequences SEQ ID No. 36 to SEQ ID No. 38 or SEQ ID No. 40 to SEQ ID No. 52) .

10

The sequences SEQ ID No. 36 to SEQ ID No. 38 and/or SEQ ID No. 40 to SEQ ID No. 52 and/or SEQ ID No. 139, used in particular in amplification reactions with sequences SEQ ID No. 34, SEQ ID No. 35 and/or SEQ ID No. 39, make 15 it possible, respectively, to detect the microorganisms

and species listed below. One or more pair(s) of sequences may be used in an amplification reaction.

Thus, the sequences according to the present invention  
5 make it possible in particular to detect microorganisms  
of the following genera and families: Lactococcus (SEQ  
ID No. 39), Bifidibacterium (SEQ ID No. 38 and/or 40),  
Mycobacterium (SEQ ID No. 40), Helicobacter (SEQ ID No.  
41), Campylobacter (SEQ ID No. 42), Bacteroides (SEQ ID  
10 No. 43), Chlamydia (SEQ ID No. 44), Mycoplasma (SEQ ID  
No. 45), Staphylococcus (SEQ ID No. 46), Lactococcus  
and/or Streptococcus (SEQ ID No. 47), Lactobacillus  
and/or Bacillus (SEQ ID No. 48), Clostridium (SEQ ID  
No. 37 and/or 49), Enterobacteriaceae (SEQ ID No. 36  
15 and/or 50), Pasteurella and/or Haemophilus (SEQ ID No.  
50), Neisseria and/or Legionella (SEQ ID No. 51),  
Aeromonas and/or Bordetella (SEQ ID No. 52),  
Lactobacillus and/or Bacillus (SEQ ID No. 139).

20 The subject of the invention is also the genomic  
sequences of microorganisms which can be amplified  
using the primers according to the invention, in  
particular the pairs of primers SEQ ID No. 32/ SEQ ID  
No. 33, and (SEQ ID No. 34, SEQ ID No. 35 or SEQ ID No.  
25 39)/(a sequence chosen from the sequences SEQ ID No. 36

to SEQ ID No. 38, SEQ ID No. 40 to SEQ ID No. 52 or SEQ ID No. 139).

Thus, a subject of the invention is also in particular  
5 a sequence from SEQ ID No. 140 to SEQ ID No. 189, which  
correspond to the hypervariable intergenic regions of  
the GroESL operon of various organisms. A subject of  
the invention is also any fragment of a minimum of 20  
bases, preferably 30 bases, more preferably 50 bases,  
10 even more preferably 75 bases, most preferably 100  
bases of one of the sequences SEQ ID No. 140 to SEQ ID  
No. 189, or the sequences complementary thereto, it  
being possible for said fragment to be used to define  
organism-specific primers, or for the identification of  
15 organisms, in particular by hybridization.

Thus, the DNA chip according to the invention  
preferably has, at its surface, a plurality of  
oligonucleotides (a minimum of two) comprising  
20 fragments chosen from the fragments of the sequences  
SEQ ID No. 140 to SEQ ID No. 189 defined above, thus  
allowing the identification of the microorganisms. The  
length of these oligonucleotides can be determined by  
those skilled in the art, as a function of the  
25 hybridization conditions, which they intend to use.

Oligonucleotides approximately 50 bases long are thus envisioned.

Subjects of the present invention are also diagnostic  
5 kits for carrying out the method according to the  
invention. These diagnostic kits contain degenerate  
primers for amplifying one or more intergenic regions  
of operon which is conserved among species. They may  
also contain the reagents required for the  
10 amplification reaction. Moreover, DNA representing  
positive or negative controls may be included in the  
diagnostic kits according to the invention.

A diagnostic kit according to the invention also  
15 advantageously contains the elements required for  
analyzing the amplified products. In particular, a  
diagnostic kit according to the invention contains a  
DNA chip according to the invention, which has, at its  
surface, the sequences corresponding to the various  
20 microorganisms.

Depending on the species it is desired to detect, the  
diagnostic kit according to the invention contains the  
appropriate pair of primers and analytical elements.  
25 Furthermore, a kit according to the invention may also

comprise instructions for carrying out the method according to the invention.

A diagnostic kit according to the invention may also  
5 only contain a DNA chip according to the invention, and  
optionally also instructions for carrying out the  
analysis of fragments located in the intergenic region  
of operons which are conserved between species, the  
preferred operons being GroESL and rpoBC.

10

The coupling of specific primers and probes thus allows  
rapid and precise identification of the flora of a  
given individual. It is therefore possible to establish  
the profile(s) of populations characteristic of healthy  
15 individuals. It is also possible to establish the  
standard profiles for various pathological conditions.

The method according to the invention also provides the  
possibility of easy monitoring of the evolution of the  
20 flora as a function of diet. More specifically, it is  
possible to follow the effects, in the colon, of a  
particular food, such as a pre- or probiotic, or of a  
medicinal treatment such as a treatment with  
antibiotics.

It is therefore possible to envision the development of foods or medicinal products for "effect on the flora" purposes, allowing reestablishment and a return to a normal profile after an imbalance subsequent to any pathological condition or attack. It is also possible to use primers and probes corresponding to pathogenic strains in order to optionally establish critical population thresholds preceding a pathological condition. It is then possible to determine which of the other populations are liable to exert a barrier effect on these pathogens.

The tools for diagnosing the intestinal flora developed and based on the method according to the invention (which are also subjects of the invention) are of interest firstly to industrial companies in the agrofoods and pharmacy domains, in order to develop their products and to determine the impact thereof on the intestinal flora. Specifically, particular diets are liable, in the long term, to significantly modify the composition of the flora and, consequently, to have harmful or beneficial effects depending on the types of population which appear or disappear. Similarly, medicinal treatments, and in particular antibiotic treatments, lead to imbalances in the microflora. Characterization of the populations affected according

to the type of medicinal product would make it possible to set up a parallel or subsequent treatment capable of preventing these modifications or of reestablishing a correct flora as rapidly as possible.

5

These tools are also of interest to health professionals, for characterizing the intestinal flora of patients, which may make it possible, for example, to direct a treatment. Specifically, gastro-  
10 enterologists estimate that 70% of the population of industrialized countries complain of diverse digestive disorders, which are called functional colopathy, ranging from simple digestive disorders such as bloating or flatulence, to more significant disorders  
15 such as constipation or diarrhea, etc. The majority of their consultations concern this functional colopathy.

Few solutions are provided to treat these disorders since, for certain types of colopathy, their cause is  
20 still quite unknown, and for others, there is no effective treatment. Added to this is the problem of the medical diagnosis since the patients presenting these symptoms of functional colopathy do not generally present any physical lesion in the colon. Only a  
25 questionnaire enables the gastroenterologist to turn toward a type of treatment, which proves to be

relatively ineffective in the majority of cases. The market for products which can relieve these disorders is therefore considerable, as is that for diagnosis. Specifically, a diagnosis of the state of the flora of the patients might provide the physician with information regarding the causes of their disorders and the treatment to be carried out.

In order to select a genomic target of interest, namely a target which is conserved in all the genomes, the conservation of the most conserved operons in the course of evolution was studied based on the genomes of the 51 bacteria entirely sequenced and available on the NCBI server. These sequences were positioned relative to that of rpoB/C (beta operon).

It emerged from this first analysis that the longest and most conserved targets are in fact the GroESL operons (encoding Hsp10 (groES) and Hsp60 (groEL)) and a part of the beta operon corresponding to the rpoB and rpoC genes (encoding the beta and beta' subunits of DNA-directed RNA polymerase). In addition, it was possible to identify conserved protein motifs sufficiently long to allow the definition and then the synthesis of universal (ubiquitous), or almost universal, degenerate primers.

Thus, these two operons were chosen in order to exemplify the principle of the method according to the invention.

5

Finally, the region of interest of the beta operon was amplified, i.e. the region amplifiable by PCR, using the two corresponding degenerate primers (FO and RP: SEQ ID No. 53 and SEQ ID No. 54) for selection of  
10 bacteria in order to establish the sequence thereof and to test them by hybridization on a nylon membrane so as to validate this specificity. These sequences were also aligned to their homologs available on GenBank in order to observe this specificity by bioinformatics.

15

The same experiments were carried with the GroESL operon, and it can thus be shown that the method according to the invention makes it possible to identify and discriminate between the various species  
20 of microorganisms.

The following examples are intended to illustrate the invention, and should not be considered as limiting the invention.

25 In the application, the abbreviations for the bacteria are as follows:

Bacillus subtilis (BS) CIP 52-65T; Bacteroides vulgatus (BV) DSM 1447; Bifidobacterium longum (BL) DSM 20219; Clostridium leptum (CL) DSM 753; Clostridium nexile (CN) DSM 1787; Clostridium spiroforme (CS) DSM 1552; 5 Clostridium glycolycum (CG) DSM 1288; Lactobacillus gasei (LG) DSM 20077; Lactobacillus helveticus (LH) CIP 103146; Lactobacillus paracasei (LP) DSM 8741; Lactobacillus reuteri (LR) DSM 20053; Pseudomonas aeruginosa (PA) CIP100720; Ruminococcus 10 hydrogenotrophicus (RH) DSM 10507; Citrobacter freundii (CF); Serratia liquefaciens (SL); Serratia marcescens (SM); Enterobacter cloacae (EnC); Escherichia coli (EsC); Morganella morganii (MM); Proteus mirabilis (PM); Klebsiella oxytoca (KO); Klebsiella pneumoniae 15 (KP).

#### DESCRIPTION OF THE FIGURES

Figure 1: Diagram of the rpoBC operon of *E. coli*. The universal (ubiquitous) primers are used to amplify the 20 intergenic sequence.

Figure 2: Diagram of the groESL operon of *E. coli*. The universal primers are used to amplify the intergenic sequence.

Figure 3: Principle of a DNA chip. Specific sequences 25 are attached to a solid support. The possible

hybridization of the complementary sequences makes it possible to determine their presence in a sample.

Figure 4: Hybridization of deposits of 10 ng of DNA amplified by PCR with rpoBC primers (i) and of genomic DNA (ii), with a *Serratia marcescens* probe (~ 0.25 ng/ml) (A) or a *Klebsiella oxytoca* probe (~ 1 ng/ml) (B) for 18 hours at 60°C and revelation for 30 minutes at 37°C. It is possible to observe cross hybridization of CF, SM, SL, EC and KP with the KO-DIG probe (~ 1 ng/ml) and of SL with the SM-DIG probe (~ 0.25 ng/ml).

Figure 5: Hybridization of deposits i (genomic DNA, a: 10 to 20 µg, b: 5 to 10 µg, c: 0.5 to 1 µg, d: 50 to 100 ng, e: 5 to 10 ng, f: 0.5 to 1 ng) and ii (DNA amplified by PCR with GroESL primers. a: 50 to 100 ng, b: 5 to 10 ng, c: 0.5 to 1 ng, d: 50 to 100 pg, e: 5 to 10 pg, f: 0.5 to 1 pg) with a PA-DIG probe (~ 10 ng/ml) for 18 hours at 42°C and revelation for 30 min at 37°C.

Figure 6: Hybridization of deposits i (DNA amplified by PCR with rpoBC primers: 10 ng/1 ng/100 pg) and ii (genome DNA: 1 µg/100 ng/10 ng) with an LR-DIG probe (~ 1 ng/ml) for 18 hours at 50, 55, 60 and 65°C and revelation for 30 min at 37°C.

## 25 EXAMPLES

### Example 1: Isolation of strains

In order to have a broad and representative sample of human colonic flora, it is necessary to isolate new bacterial strains, also called nonculturable (and  
5 therefore unknown) strains, which constitute a high percentage of the bacteria of the human colonic flora.

In order to perform these isolations, a large quantity of human stools is collected and sterilized by means of  
10 gamma-type radiation or by heat, for the purpose of sealing therein samples of these same human stools and of culturing them aerobically and anaerobically, in liquid and solid media.

15 Depending on the culture conditions used, it is thus possible to isolate new genera, species or strains of colonic bacteria or other eukaryotic microorganisms.

Example 2: Characterization of the sequences of the  
20 isolated strains (rpoB)

This involves carrying out the molecular characterization of sequences, ideally of mRNA to perform a quantification, and if not of genomic DNA, of  
25 the isolates of bacterial or eukaryotic microorganisms.

The sequences corresponding to portions of the bacterial rpoBC operon are studied. The genes of this operon are in fact relatively homologous between genera.

5

A computer analysis (sequence alignment) thus makes it possible to define degenerate primers for amplifying the region which is heterologous between species and which corresponds to the transcribed intergenic region.

10

Thus, the primers SEQ ID No. 1 and SEQ ID No. 31, and also the other primers, were defined after alignment of the sequences corresponding to more than 50 species of living organisms (prokaryots and eukaryots, not shown) using the redundancy of the genetic code. The sequences SEQ ID No. 53 and SEQ ID No. 54 are in particular preferred.

15

The regions amplified by PCR or RT-PCR with the abovementioned primers can obviously be cloned into various vectors, in order to be used to refine the analysis (in particular in order to sequence them).

20

Example 3: Characterization of the sequences of the isolated strains (GroESL)

25

The sequences corresponding to portions of the bicistronic bacterial GroESL operon are studied. The genes of this operon are in fact relatively homologous between genera.

5

Computer analysis (sequence alignment) thus makes it possible to define degenerate primers for amplifying the region which is heterologous between species and which corresponds to the transcribed intergenic region.

10

Thus, the primers SEQ ID No. 34 and SEQ ID No. 35 were defined after alignment of the sequences corresponding to more than 100 species of living organisms (prokaryots and eukaryots, not shown).

15

The sequences SEQ ID No. 36 to SEQ ID No. 52, and in particular SEQ ID No. 139, correspond to complementary sequences which can be used to amplify microorganisms of diverse genera and/or families.

20

As regards the primers SEQ ID No. 32 and SEQ ID No. 33, they were defined based on the conserved sequences of the GroES and GroEL genes of *E. coli*, using the degenerative genetic code.

25

Example 4: Amplification reactions (GroESL)

The PCR reactions are carried out according to the following protocol:

5 2 ml of culture broth shaken at 37°C for 18 h are concentrated by centrifugation and resuspension of the bacterial pellet in 30 µl of distilled water, and then a 1/10 dilution of this concentrate, treated at 100°C for 10 minutes, is used as a matrix for the PCR  
10 reactions. The reaction conditions are 94°C/5 min, then 25 cycles of (94°C/30 sec, 60°C/45 sec, 72°C/30 sec), followed by an elongation step at 72°C for 7 min.

Analysis of the amplicates makes it possible to show  
15 that it is possible to amplify, using the primers SEQ ID No. 32 and SEQ ID No. 33, the intergenic region of various enterobacteria, such as *Escherichia coli*, *Enterobacter cloacae*, *Morganella morganii*, *Serratia liquefaciens*, *Proteus mirabilis*, *Serratia marcescens*,  
20 *Klebsiella pneumoniae*, *Citrobacter freundii* or *Klebsiella oxytoca*. The amplified region varies in length, according to the species, from 400 to 500 base pairs (bp). Use of the pair SEQ ID No. 34 and SEQ ID No. 36 gives amplicates of between 550 and 650 bp in  
25 length.

Use of the pairs: (SEQ ID No. 34, SEQ ID No. 35 or SEQ ID No.39)/(a sequence chosen from the sequences SEQ ID No. 36 to SEQ ID No. 38 or SEQ ID No. 40 to SEQ ID No. 52, or SEQ ID No. 139) makes it possible to amplify  
5 sequences specific to certain families and species, and to identify the organisms of these families or species.

For the amplification reactions, use is preferably made of a primer marked "A" with a primer marked "B".

10

The regions amplified by PCR or RT-PCR with the abovementioned primers can obviously be cloned into various vectors, in order to be used to refine the analysis (in particular in order to sequence them).

15

#### PCR PROTOCOL

In order to show that using the intergenic region of the two operons of interest as a nucleic acid probe can  
20 make it possible to discriminate several bacterial species, said IGR for each target was amplified by direct PCR on bacterial suspensions. For the amplification reactions, use is preferably made of a primer marked "A" with a primer marked "B".

25

2 ml of culture broth shaken at 37°C for 18 h are concentrated by centrifugation and resuspension of the bacterial pellet in 30 µl of distilled water, and then a 1/10 dilution of this concentrate, treated at 100°C for 10 minutes, is used as a matrix for the PCR reactions.

groESL OPERON:

The PCR reactions for this target are carried out at a  $T_m$  ranging between 59°C and 60°C. The reaction conditions are 94°C/5 min, then 25 cycles of (94°C/30 sec, 60°C/45 sec, 72°C/30 sec), followed by an elongation step at 72°C for 7 min. The amplified intergenic regions are then observed by agarose gel electrophoresis using a 1 Kb + ladder (Gibco BRL).

Analysis of the amplicates makes it possible to show that it is possible to amplify, using the primers SEQ ID No. 32 and SEQ ID No. 33, the intergenic region of various Enterobacteria, such as *Escherichia coli*, *Enterobacter cloacae*, *Morganella morganii*, *Serratia liquefaciens*, *Proteus mirabilis*, *Serratia marcescens*, *Klebsiella pneumoniae*, *Citrobacter freundii* or *Klebsiella oxytoca*. The amplified region varies in length, according to species, from 400 to 500 base

pairs (bp). Use of the pair SEQ ID No. 34 and SEQ ID No. 36 gives amplicates of between 550 and 650 bp in length.

5 rpoB/C OPERON:

The PCR reactions for this target are carried out at a  $T_m$  ranging between 63°C and 64°C. The reaction conditions are 94°C/4 min, then 30 cycles of  
10 (94°C/30 sec, 64°C/30 sec, 72°C/3 min), followed by an elongation step at 72°C for 12 min. The amplified intergenic regions are then observed by agarose gel electrophoresis using a molecular weight marker III DNA ladder (ref: No. 528552; Boehringer Mannheim).

15

Analysis of the amplicates makes it possible to show that it is possible to amplify, using the pair of primers SEQ ID No. 53 and SEQ ID No. 54, the intergenic region of the various bacteria, such as *Escherichia*  
20 *coli*, *Clostridium leptum*, *Klebsellia oxytoca*, *Lactococcus lactis*, *Citrobacter freundii*, *Serratia marcescens*, *Proteus mirabilis*, *Serratia liquefaciens*, *Morganella morganii*, *Enterobacter cloacae* or *Ruminococcus hydrogenotrophicus*.

25

DNA fragments corresponding to the intergenic regions of the rpoB/C operon in various species were reamplified and analyzed using bands extracted from an agarose gel preparation. These fragments were  
5 prepurified with a Qiagen extraction kit.

The regions amplified by PCR or RT-PCR with the abovementioned primers can obviously be cloned into various vectors, in order to be used to refine the  
10 analysis (in particular in order to sequence them).

#### HYBRIDIZATION PROTOCOL

With a view to testing the specificity of the PCR  
15 products, for the species selected for our study, deposits of these DNAs were made on a nylon membrane according to a sodium hydroxide (NaOH) fixation protocol. The DNA concentrations for these deposits are given on the corresponding figures. These membranes  
20 were hybridized according to the protocol of the PCR DIG Probe Synthesis Kit (Roche) Cat. No. 1636090. The concentration of the probe used, synthesized according to the same protocol, is also indicated on the figures, as is the temperature of hybridization carried out  
25 overnight (18 h). The temperature of pre-hybridization is 65°C for each experiment, and it lasts 45 min.

Detection of this type of hybridization with this type of labeling (DIG) is termed colorimetric ("cold" labeling different from radioactive labeling).

5

Figures 4 to 6 show a specificity detection as a function of the organisms, although some crosshybridization reactions may exist. These reactions may be reduced by choosing probes which are shorter and  
10 located among the hypervariable intergenic sequences, as defined by SEQ ID No. 63 to SEQ ID No. 138 (rpoN) or SEQ ID No. 40 to SEQ ID No. 189 (GroESL).

Thus, a DNA chip with various probes located in the  
15 intergenic region will make it possible to recognize without hesitation the presence or absence of a microorganism, even when there is crosshybridization. Specifically, the presence of a microorganism will be deduced from the hybridization for each of the probes.

20

It may therefore be advantageous to define DNA chips having specific probes corresponding to the intergenic region of each microorganism, but also to include several different probes for each microorganism.

## References

- {1} - Welling, et al., Applied Environmental Microbiology, **64**:3336-45, 1998: Variations of bacterial  
5 populations in human feces measured by fluorescent in situ hybridization with group-specific 16S rRNA-targeted oligonucleotide probes.
- {2} - Greisen, et al., Journal of Clinical Microbiology, **32**:335-351, 1994: Staphylococcal  
10 identification using PCR amplification of 16S rDNA genes.
- {3} - Jensen, et al., Applied Environmental Microbiology, **59**:945-952, 1994: Staphylococcal  
15 identification using PCR amplification of spacer regions between 16S and 23S genetic loci.
- {4} - Plikaytis, et al., above; Telenti, et al., above: PCR strategies for Mycobacterial speciation requiring the detection of restriction site polymorphisms (RFLP) within amplified products.
- 20 {5} - Zoetendal, Akkermans and De Vos, Applied Environmental Microbiology, **64**:3854-9 (1998): Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria.
- 25 {6} - Vaitilinggom, Gendre and Brignon, Applied Environmental Microbiology **64**:1157-60, 1998: Direct

detection of viable bacteria, molds, and yeast by reverse transcriptase PCR in contaminated milk samples after heat treatment.

{7} - Gendre and Brignon, Compagnie Gervais Danone,  
5 patent # PCT/FR97/01918 WO 98/18958 (1998): Method for detecting live microbiological contaminants in a food product sample.

{8} - Doré, et al., Applied Environmental Microbiology, 65:4799-807, 1999: Direct analysis of  
10 genes encoding 16S RNA from complex communities reveals many novel molecular species within human gut.

{9} - Goh, Chow and Hemmingsen, US patents 5,708,160 (1998) & 5,989,821 (1999).

{10} - Emelyanov and Sinitsyn, Russian Journal of Genetics, 35:618-627, 1999: A GroE-based phylogenetic  
15 analysis shows a very close evolutionary relationship between mitochondria and rickettsia.